ANGLE MODULATION

In amplitude modulation, we saw that we could impose the message signal upon the carrier signal by varying the amplitude of the carrier. The phase and frequency of the carrier were left unchanged. In angle modulation the amplitude is left unchanged but either the phase or frequency of the carrier is varied in a manner proportional to the message signal.

If we start with the normal carrier frequency at f_c, $E_c\cos(2\pi f_c t)$, the phase of the waveform is $\theta(t) = \theta_c(t) = 2\pi f_c t$. If we now modify $\theta(t)$ linearly with the message signal, i.e., $\theta(t) = \theta_c(t) + k_p a(t)$ where k_p is the phase sensitivity (in radians per volt or radians per ampere) we will have a phase modulated waveform. Our phase modulated waveform will be

$$m(t) = E_c\cos(\theta_c(t) + k_p a(t)) = E_c\cos(2\pi f_c t + k_p a(t)). \tag{1}$$

A frequency modulated waveform takes a little different form. If we look first at a sinusoidal waveform of unknown frequency, we could characterize it as

$$m(t) = E\cos(2\pi f t) = E\cos(\theta(t)). \tag{2}$$

To determine the frequency of this signal, we can differentiate the phase, $\theta(t)$, to get

$$\frac{d\,\theta(t)}{dt} = 2\pi f \qquad \Rightarrow \qquad f = \frac{1}{2\pi}\frac{d\,\theta(t)}{dt}. \tag{3}$$

Knowing that the phase of the carrier wave is $\theta_c(t) = 2\pi f_c t$, if we can modify the frequency of $\theta_c(t)$ such that the phase, $\theta(t)$, of a frequency modulated wave is no longer a function of just f_c, but a function of the instantaneous frequency, $f_i = f_c + k_f a(t)$. The constant k_f is the frequency sensitivity in hertz per volt (or hertz per ampere). By combining with the above equation we can see that

$$f_i = f_c + k_f\, a(t) = \frac{1}{2\pi}\frac{d\ \theta(t)}{dt}. \tag{4}$$

To find $\theta(t)$ we simply integrate the quantity $2\pi f_i$, so that

$$\theta(t) = 2\pi\int_0^t f_i dt = 2\pi\int_0^t [f_c + k_f\, a(t)]dt = 2\pi f_c t + 2\pi k_f \int_0^t a(t)dt. \tag{5}$$

The frequency modulated waveform is therefore

$$s(t) = E_c \cos\left(2\pi f_c t + 2\pi k_f \int_0^t a(t)dt\right). \tag{6}$$

Comparing Equations 1 and 6 we can see that they are of the same general form and the frequency modulated waveform is simply a phase modulated waveform with the integral of a(t) modifying the phase rather than a(t) modifying it as in Eq. 1. Similarly, a phase modulated waveform is just a frequency modulated waveform with the differentiated integral of a(t) modifying the waveform rather than the integral.

Given this relationship it suffices to analyze one or the other and the one not analyzed can be inferred from the other. Therefore we will only analyze Frequency Modulation (FM).

A. FM MODULATION

In the last section we saw that the instantaneous frequency of a frequency modulated wave is

$$f_i = f_c + k_f a(t) \tag{7}$$

so that the FM wave could be described by

$$m(t) = E_{c_c} \cos\ (2\pi f_c t + 2\pi k_f \int_0^t a(t)dt). \tag{8}$$

To begin analysis of this modulation form let's start with a sinusoidal modulating signal of constant amplitude, i.e.,

$$a(t) = E_m \cos(2\pi f_m t). \tag{9}$$

the instantaneous frequency of the FM wave will then be

$$f_i = f_c + k_f a(t) = f_c + k_f A_m \cos(2\pi f_m t). \tag{10}$$

Just as we let m represent the percent of modulation in AM, we define a term in FM called the frequency deviation, Δf. The frequency deviation is defined to be the frequency sensitivity, k_f, multiplied by the maximum amplitude of the modulating signal and is a measure of how far the instantaneous frequency will deviate or change from the carrier frequency. In the case of the sinusoidal modulating signal, $\Delta f = k_f E_m$, so that our instantaneous frequency can be defined as

$$f_i = f_c + \Delta f \cos(2\pi f_m t). \tag{11}$$

Our FM waveform can therefore be defined as

$$\begin{aligned} m(t) &= E_c \cos\left(2\pi f_c t + 2\pi k_f \int_0^t a(t)dt\right) \\ &= E_c \cos\left(2\pi f_c t + 2\pi \Delta f \int_0^t \cos(2\pi f_m t)dt\right) \\ &= E_c \cos\left(2\pi f_c t + \frac{\Delta f}{f_m}\sin(2\pi f_m t)\right). \end{aligned} \tag{12}$$

We now define the modulation index for FM as $\Delta f/f_m$ and we will call it β. As we will soon see, this will be the variable for a Bessel function, i.e., we will use β in Bessel function calculations. Using this notation, the FM signal is defined as

$$m(t) = A_c \cos(2\pi f_c t + \beta \sin(2\pi f_m t)). \tag{13}$$

(Note book uses m instead of β.)

The following part is for completeness only. You are not responsible for it.

We can find the in-phase and quadrature components $m_I(t)$ and $m_Q(t)$ of this signal using the trig identity

$$m(t) = E_c \cos(2\pi f_c t)\cos(\beta \sin(2\pi f_m t)) - E_c \sin(2\pi f_c t)\sin(\beta \sin(2\pi f_m t)), \tag{14}$$

so that $m_I(t) = E_c \cos(\beta \sin(2\pi f_m t))$ and $m_Q(t) = E_c \sin(\beta \sin(2\pi f_m t))$.

Using the techniques of the complex envelope, we know that

$$\tilde{m}(t) = m_I(t) + j m_Q(t). \tag{15}$$

Substituting our values for the in-phase and quadrature terms, we get

$$\begin{aligned} \tilde{m}(t) &= E_c[\cos(\beta\sin(2\pi f_m t)) + jsin(\beta\sin(2\pi f_m t))] \\ &= E_c e^{j\beta\sin(2\pi f_m t)}. \end{aligned} \tag{16}$$

It is clear that m(t) is periodic so that we can define it in a Fourier series, to give

$$\tilde{m}(t) = \sum_{n=-\infty}^{\infty} c_n e^{j2\pi n f_m t} \tag{17}$$

where

$$\begin{aligned} c_n &= \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} \tilde{m}(t)\ e^{-j2\pi nt/T_0}\ dt \\ &= f_m \int_{-1/2f_m}^{1/2f_m} \tilde{m}(t) e^{-j2\pi n f_m t} dt \\ &= f_m E_c \int_{-1/2f_m}^{1/2f_m} e^{j\beta\sin(2\pi f_m t) - j2\pi n f_m t} dt. \end{aligned} \tag{18}$$

Let $x = 2\pi f_m t$, then $dx = 2\pi f_m dt$ and the limits of integration will be $-\pi$ to π so that

$$c_n = \frac{E_c}{2\pi}\int_{-\pi}^{\pi} e^{j(\beta\sin x - nx)} dx = E_c J_n(\beta). \tag{19}$$

This result allows us to rewrite the equation for the complex envelope as

$$\tilde{m}(t) = E_c \sum_{n=-\infty}^{\infty} J_n(\beta) e^{j2\pi n f_m t} \tag{20}$$

and recalling the relationship between the complex envelope, the pre-envelope, and m(t), we know that m(t) can be found from the complex envelope by shifting in frequency and taking the real part, i.e., (respite over)

$$\begin{aligned} s(t) &= A_c \ Re\left[\sum_{n=-\infty}^{\infty} J_n(\beta) e^{j2\pi(f_c+nf_m)t}\right] \\ &= A_c \sum_{n=-\infty}^{\infty} J_n(\beta)\cos(2\pi(f_c + nf_m)t). \end{aligned} \tag{21}$$

Upon inspection, we can see that our signal is composed of a series of constants ($E_c\ J_n(\beta)$) multiplied by cosine waves of frequencies, f_c (n=0), $f_c \pm f_m$ (n= ± 1), $f_c \pm 2f_m$, etc., where the positive components make up the upper sidebands and the negative the lower sidebands. Theoretically, the sum consists of the infinite harmonics of f_m.

We can define the FM waves created in this fashion as either Narrow-band or Wide-band. This definition arises from a comparison with AM. Recall that for tone modulation of an AM carrier we had the carrier and an upper sideband at f_c+f_m and a lower sideband at f_c-f_m. If we define narrow-band FM to have this same characteristic, we can see that the infinite summation above is reduced to having Bessel function magnitudes above zero only for n=0 and n= ± 1, i.e., $J_n(\beta)=0$, $|n|>1$. We find this is true for $\beta \leq 0.3$, where $J_0(\beta)\approx 1$, and $J_1(\beta)\approx\beta/2$ (see Table 9.1 p. 276). Substituting these values we find that the narrow-band FM signal is

$$m(t) \approx E_c\cos(2\pi f_c t) + \tfrac{1}{2}\beta E_c\cos(2\pi(f_c + f_m)t) - \tfrac{1}{2}\beta E_c\cos(2\pi(f_c - f_m)t), \tag{22}$$

which is composed of the carrier and the two sidebands. This looks very much like the AM signal except that the lower sideband can be seen to be negative in this case.

For wide-band FM, β is not constrained and the sidebands consist of the infinite frequency harmonics of f_m. The magnitude of the components is controlled by the modulation index, β, and therefore $J_n(\beta)$.

How many of these sidebands are important for the transmission of the FM signal? Another way to state this is how much bandwidth do we require to adequately transmit the signal? We stated that the frequency deviation, Δf, defined the amount of deviation of the modulating signal away from the carrier frequency. But, because of the infinite summation of the sidebands, the sideband frequencies will exceed Δf, but we can see from Table 9.1 that the magnitude rapidly approaches zero for those sidebands above Δf. Therefore, the bandwidth of the signal, W, always exceeds Δf, but is limited.

In trying to define the bandwidth, a fellow named J.R. Carson in the 1920s noticed that for large β, the bandwidth is approximately equal to $2\Delta f$. However, for small β the bandwidth was closer to $2\,f_m$, for a single tone, or 2 W in general. He proposed a bandwidth definition which is still used today called Carson's rule which is

$$B \approx 2\Delta f + 2W = 2(\Delta f + W). \tag{23}$$

However, Carson's rule generally underestimates the bandwidth requirement of the signal. A better estimate is one called Carlson's rule which allows for more of the spectral lines. Carlson's rule is

$$B \approx 2(\Delta f + 2W). \tag{24}$$

The power contained in a sinusoid of constant amplitude, E_c, is constant as well, i.e., $P = ½ E_c^2$. This power relationship is valid regardless of the frequency of the sinusoid, so long as the amplitude remains constant. Therefore, the average power of the FM wave remains constant at $½ E_c^2$.

B. FM DEMODULATION

Now that we have modulated the frequency of our carrier, how do we recover the original message signal? Since the message is modulated within the signal, just as we did with AM, we must demodulate to recover the message, a(t). However, the methods of AM, i.e., envelope detection (without some pre-detection) and coherent demodulation, will not work with FM. FM demodulators are often called discriminators and work on different principles than we have seen before. We can demodulate by a direct or by an indirect method. We will look at a method of direct demodulation.

The most prevalent type of direct method of demodulation is called the Balanced Frequency Discriminator. To understand the frequency discriminator, recall the equation for the FM wave,

$$m(t) = E_c \cos\left[2\pi f_c t + 2\pi k_f \int_0^t a(t)dt\right]. \tag{25}$$

If we differentiate m(t) with respect to time we get

$$m'(t) = -E_c[2\pi f_c + 2\pi k_f\ a(t)]\ \sin\left[2\pi f_c t + 2\pi k_f \int_0^t a(t)dt\right]$$
$$= -2\pi f_c E_c \left[1 + \frac{k_f}{f_c} a(t)\right] \sin[\ldots]. \tag{26}$$

We can see that the envelope of m'(t) is $2\pi f_c E_c[1 + k_f/f_c\ a(t)]$ which can be demodulated with an envelope detector. Therefore, to demodulate the FM signal we only need to differentiate it and send it through an envelope detector.

You may notice that the effective carrier frequency of Eq. 26 above is not the same as f_c but deviates around it. As long as f_c is much greater than f_m the detector will not follow these changes in frequency and no distortion will occur. However, distortion can occur with this demodulator, due to non-linearities in the two filters.